Risk of Death for Low Birth Weight Babies Using Cox Regression

Dewi Murni[1], Yenni Kurniawati[2], Rosi Ramayanti[3]
{dewimurni_mat@fmipa.unp.ac.id[1]}

Lecturers of Mathematics Department State University of Padang, Indonesia[1,2], Student of Mathematics Department State University of Padang, Indonesia[3]

Abstract. Low Birth Weight (LBW) is one of the factors triggering the occurrence of infant mortality in the neonatal period. Factors affecting the risk of low birth weight is maternal death, fetal, infant birth weight, and socio-economic factors. That this study aims to look at the factors that affect the risk of death LBW base on the chance of death factors. One statistical analysis used in looking at the relationship of LBW risk of death is the Cox regression model. The results of this study indicate that the factors that influence the risk of LBW death are gestational age and infant birth weight. Opportunity risk of death for LBW who have gestational age < 37 weeks is 8.424 times greater than the LBW who have ≥ 37 weeks gestation. Then chances LBW risk of death with birth weight < 1499 gr was 8.547 times greater than the LBW with birth weight $1500 - 2499$ gr.

Keywords: LBW, Neonatal, Cox regression

1 Introduction

Infant Mortality Rate (IMR) is one of the important indicators in determining the level of public health. The success of health development in an area can also be seen from the Infant Mortality Rate (IMR) and Life Expectancy Rate (LER). Most of these IMRs are deaths during the neonatal period [1]. According to the Indonesian Demographic and Health Survey data in 2007 (2007 IDHS), the neonatal mortality rate in Indonesia is 19 deaths / 1000 live births and IMR is 34 deaths / 1000 live births. However, in 2009 AKB 26/1000 live births. This figure is still far from the Millennium Development Goals (MDGs) target that must be achieved in 2015, namely IMR 19 / 1,000 KH. Therefore, this, of course, must be followed up with efforts to accelerate (accelerate) the reduction in maternal and newborn mortality rates.

One risk factor that has contributed to neonatal mortality is the incidence of Low Birth Weight Babies (LBW). LBW refers to a birth weight of 500 - 2500 grams. LBW is 35 times higher risk of death compared to babies whose birth weight is above 2500 grams [2].

Some risk factors that influence the risk of LBW death are maternal factors, fetal factors, infant birth weight factors and socioeconomic factors [3]. Many factors are suspected as factors that influence the risk of LBW death. Many factors are suspected as factors that influence the risk of LBW death. To determine these factors analysis is needed. In this study, the dependent variable (Y) is the LBW endurance time. Endurance time tends to have skewness [4]. Because more data is to the left of the average, this means that more data is less than the average value. This is because the LBW endurance time tends to be short. After all, the risk of failure is higher.

As a result, the tail of the distribution function is longer to the right interval and the distribution becomes asymmetrical. Therefore, in general, the holding time is not normally distributed.

There is an analysis that is used to see the relationship between the durability of an object with several explanatory variables, the analysis is proportional hazard analysis or known as cox regression. Analysis of the relationship can be stated in a Cox regression model or proportional hazard model. The Cox regression model is a statistical technique to see the relationship between the survival of an object with several explanatory variables [5]. Through this cox regression model, the risk of failure can be seen at a certain time, wherein this study the failure in question is that LBW is unable to survive or is declared dead. Cox regression models in the form of hazard functions for i-individuals where i = 1,2, ..., n, named Equation 1.

$$h_i(t) = exp\big(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}\big)h_0(t). \tag{1}$$

Where $h_0(t)$ is a hazard function obtained from the opportunity-dense function of the Weibull distribution, for example, see Equation 2.

$$h_0(t) = \lambda \gamma t^{\gamma - 1} \tag{2}$$

The estimation of parameters in this cox regression model can be solved using the maximum likelihood method. The significance test of the model used in this analysis is the likelihood ratio test, with test statistics $\chi^2_{LR} = 2[l(\hat{b}) - l(b_0)]$. Test the significance of the parameters in the model resulting from the estimation is used the Wald test, with the test statistic $W^2 = \left[\frac{\beta_i}{SE(\beta_i)}\right]^2$ if $W^2 > \chi^2_{\alpha,1}$ this means that explanatory variables have a contribution to the model and choose the best model using the Akaike's Information Criterion (AIC) method [4].

2 Method

This research is applied statistics. The method used is descriptive method with analysis of LBW medical record data at Padang Panjang Hospital in 2012-2013 with a population of 130 LBW and a sample of 71 LBW people. In carrying out this research, the author begins by reviewing the problem, gathering reference material, linking the problem with the statistical theory obtained with the problem so that it can answer the problem and finally drawing conclusions from the problems that have been discussed. The work steps to be carried out are recapitulating LBW medical record data at Padang Panjang Hospital, describing data, creating dummy variables, testing data distribution by survival time data with the help of Minitab software, forming cumulative functions that correspond to data distribution, forming survival function, establish hazard function, test proportional hazard modeling assumptions through $\ln[-\ln \widehat{S}(t)]$, estimating distribution parameters using SAS software, estimating model parameters using SAS software, testing the significance of the model using SAS software, examining the role of parameters in the model with the Wald test, selecting the best model using the AIC method, interpreting and drawing conclusions.

3 Results and Discussion

3.1 Research Results

Data description. Data on medical records of the number of LBW undergoing hospitalization at the Padang Panjang Hospital based on a bar diagram of several factors that influence the risk of LBW death is illustrated in the **Figure 1** below. The risk factors for death are maternal age which consists of 3 categories namely <20 years, 20-35 years and> 35 years. The maternal height factor consists of 2 categories: mother's height <150 cm and ≥150 cm. Maternal weight factors consist of 2 categories: <50 kg and total parity of ≥50 kg. The parity factor or the number of children born has two categories, namely the number of parities <4 people and ≥4 people. Factors of history of the maternal disease which consists of 2 categories, namely mothers who have a history of diseases such as DM, hypertension anemia and no such disease. Baby sex factors consist of 2 categories, namely male and female. Gestational age factors consist of 2 categories, namely gestational age <37 weeks and ≥37 weeks. Infant birth weight factors consist of 3 categories, namely <1000 gr, 1500-2499 gr, and 1000-1499 gr.

Maternal education factors consist of 3 categories, namely graduating from junior high school, graduating high school, and graduating from University. Mother occupational factors consist of 3 categories, namely housewives, civil servants, and the private sector.
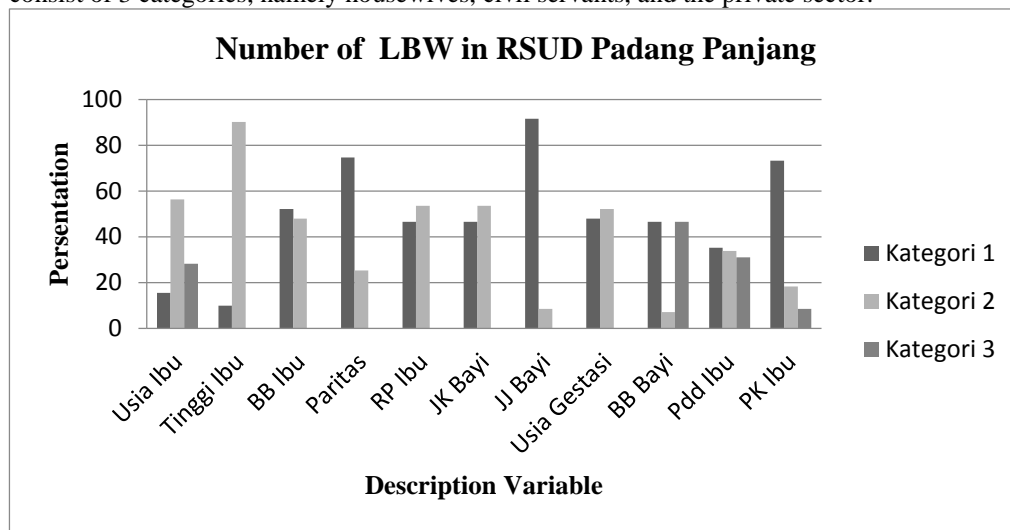
Fig. 1. Number of LBW in RSUD Padang Panjang.

Based on the diagram above, it can be seen that maternal age from LBW in RSUD Padang Panjang is highest at 20-35 years, where the percentage reaches 56.34% and maternal age> 35 years the percentage is 28.26%, while maternal age <20 years the percentage is 15.49%. Based on the variable mother's height, it can be seen that LBW who has a mother with a height <150 cm is 9.86%, while LBW who has a mother with a height of ≥150 cm is 90.14%. Maternal weight indicates the nutritional status of the mother since childhood. The poor maternal nutritional status will interfere with the development of the fetus during the womb, causing the birth of LBW. Babies born with a mother weighing <50 kg have twice the risk compared to

babies born with a mother weighing ≥50 kg [2]. This is consistent with what happened to LBW in Padang Panjang Hospital. As shown in Table 3, LBW has a mother with a bodyweight <50 kg more than LBW who has a mother with a weight body ≥of 50 kg, where the percentage is 52.11% and 47.89%. Based on the parity variable, it can be seen that mothers who have parity of <4 people are 74.65%, while mothers who have parity of ≥ 4 people amount to 25.35%. One of the factors that influence the risk of LBW death is a history of maternal disease, it can be seen that mothers who have a history of diseases such as DM, anemia, and hypertension are 46.48% and for mothers who have no disease history of 53.52%. Based on the baby's sex variable, it can be seen that LBW is more female than male. Where LBW is 53.52% for women, while 46.48% for men. Based on the variable type of the fetus of the baby it can be seen that the number of LBW with a single fetus is more than that of a twin fetus, where the percentage is 91.55%, while the percentage of twins is 8.45%. Based on the gestational age variable, it can be seen that LBW with gestational age <37 weeks more than LBW with gestational age ≥37 weeks, where the percentage of gestational age <37 weeks is 52.11%, while the percentage of gestational age ≥37 weeks is 47.89%. This is consistent with those babies born with gestational age <37 weeks will have a low birth weight and have the highest risk of neonatal death [2]. Another factor influencing the risk of LBW death is the birth weight of the baby. From table 9, it can be seen that the highest birth weight of LBW is <1000 gr and 1500-2499 gr, where the percentage is 46.47%, while for birthweight LBW 1000-1499 gr the percentage is 7.06%. The level of education of a mother will affect its ability to maintain the condition of pregnancy, wiser to choose birth attendants and care for babies so that it will have an impact on the survival of the LBW [1]. This is by what happened to mothers of LBW in Padang Panjang Hospital, the highest number of LBW occurred in mothers with junior high school education, where the percentage was 35.21%. Whereas for mothers who graduated from high school and PT the percentage was 33.80% and 30.99%. When viewed in terms of employment, the occupation of mothers of LBW in RSUD Padang Panjang more RT mothers is 73.24%. For civil servants and private sector is 18.30% and 8.46%.

Data distribution test. Data distribution testing is performed on survival time data through the help of SAS software. The hypotheses in this test are:

H_0: Data follows the Weibull distribution
H_1: Data does not follow the Weibull distribution
Based on the results of SAS output in Appendix 2, it appears that the cumulative probability plot points approach a straight line, and the value of the shape parameter or $\gamma > 0$, this means there is a positive slope or skewness that causes the solid function of the opportunity to tilt to the right, then accept H_0, this means that the time of resilience is distributed to Weibull.

Analysis result. *Hazard proportional modeling assumptions test.* Before modeling several factors suspected to influence the risk of LBW death, first testing the proportional hazard modeling assumptions of each independent variable through the pattern of $ln[-ln\widehat{S}]$ such as the **Figure 2-4** below.
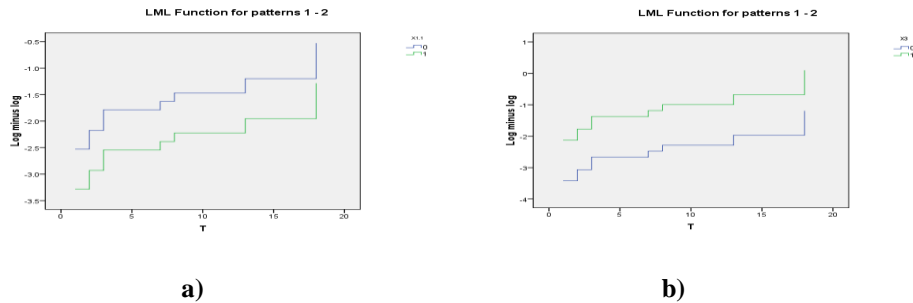
Fig. 2. a) Assumption test of the Mother's Age variable towards survival time, b) Assumptions Test of the mother's weight variable towards survival time
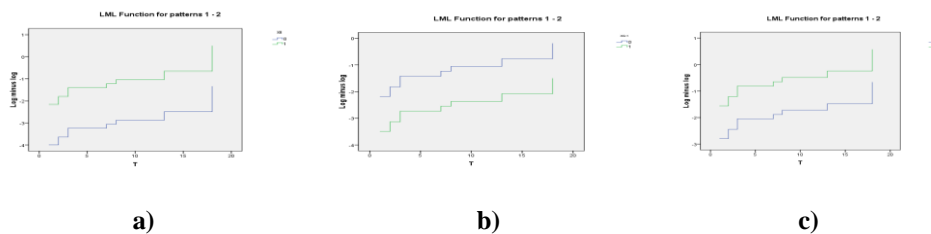
Fig. 3. a) Uji asumsi variabel Jenis Kelamin Bayi terhadap waktu ketahanan hidup, b) Assumptions Test of Baby's weight variable on survival time, c) Assumptions Test of Baby's weight variable towards survival time
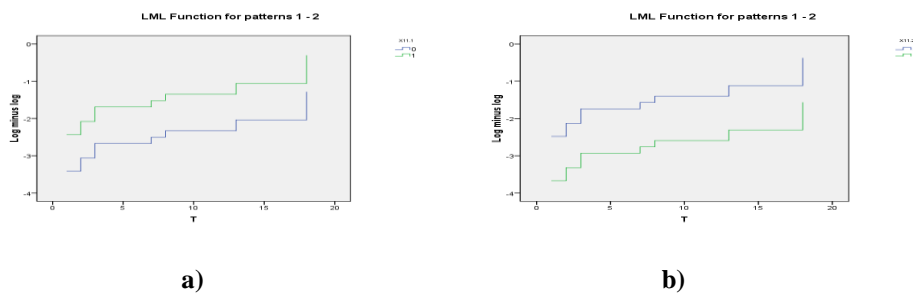
Fig. 4. a) Test the assumptions of the Mother Education variable on survival time, b) Test Assumptions for Mother Age variables on survival time

Based on the SPSS output above, it can be seen that what meets the proportional hazard assumption is maternal age, maternal weight, gestational age, infant birth weight, and maternal occupation. While other factors do not meet the assumptions, because the plots between categories in one variable do not look parallel. So for further analysis, variables that do not meet this assumption are not included.

Cox regression model. Before estimating the parameters of the model, distribution parameters will first be estimated. Since the distribution used in this study is the Weibull distribution, we will predict the parameters λ and γ of the Weibull distribution.

By using SAS software, the output as shown in Appendix 6 is obtained, from the results, the presumptions of the μ (intercept) and scale parameters are as described in Table 1.

Table 1. Estimation of distribution parameter.

Parameter	Estimate
Intercept	5.5465
Scale	0.9516
Weibull Shape	1.0508

Therefore we have Equation 3.

$$\hat{\lambda} = exp\left(\frac{-\mu}{\sigma}\right) = exp\left(\frac{-5.5465}{0.9516}\right) = 0.002942$$

$$\hat{\gamma} = \frac{1}{\sigma} = \frac{1}{0.9516} = 1.05086 \tag{3}$$

Based on SAS output, the Cox regression model in this study is in Equation 4.

$$h(t) = 0.003918(t)^{0.05086}\exp(-0.1747X_{1.1} + 0.8666X_3 - 1.7479X_8 - 1.6363X_{9.1} + 0.2955X_{9.2} + 0.4343X_{11.1} - 0.3050X_{11.2}) \tag{4}$$

a. Model Signification Test.

After the model is obtained, the next step is to test the significance of the model. The model significance test uses the likelihood ratio test, based on a full model with a null model. Using calculation from SAS software, the log-likelihood value for the null model is -51.15374236 and the log-likelihood value for the full model is -40.48374803, so the result got from χ^2_{LR} value is obtained in Equation 5.

$$\chi^2_{LR} = 2[l(\hat{b}) - l(\hat{b}_0)] = 2(-40.48374803 + 51.15374236) = 21.3399 \tag{5}$$

By using a significance level of 0.05 and free degree 5, a value of $\chi^2_{0.05,6} = 12.592.$ is obtained. Because $\chi^2_{LR} > \chi^2_{0.05:5}$, reject H_0 or accept H_1. This means that there is at least one parameter estimator that is not equal to zero. So, this model is significant

b. Model Signification Test

To find out the role or contribution of each variable to the model, a Wald test was performed. By using the equation the wald and p-values of each independent variable are obtained. From these results the p-value for X1.1 is 0.8701; X3 is 0.2083; X8 is 0.0332; X9.1 is 0.0191; X9.2 is 0.6565; X11.1 is 0.6989 and X11.2 is 0.8277. By using a real level of 0.05, the p-values of X8 and X9.1 are smaller than 0.05. While the p-values for X1.1, X3, X9.2, X11.1, and X11.2 are greater than 0.05. This means that the variables X8 and X9.1 have a significant role in the model, while the variables X1.1, X3, X11.1, and X11.2 are not significant to the model.

c. Form the best regression model

For the selection of the best model, the test procedure that will be used is the Akaike's Information Criterion (AIC) method, by comparing the full model with the reduction model (the model that only involves X8 and X9.1). Using SAS software, the log-likelihood value for the full model is -40.4837480 and the reduction model is -43.0246624. So the AIC value for each model is stated in Table 2.

Table 2. AIC value.

Model	AIC
Full Model	101.96749
Reduction Model	92.04932

From Table 2 it can be seen that the model that has the smallest AIC value is the reduction model. So the best model is a model that only involves X8 and X9.1. Using SAS software, an output is obtained, based on these results the estimated Weibull distribution parameters for the best model are in Table 3.

Table 3. Estimation of Distribution Parameter for the best model.

Parameter	Estimate
Intercept	6.3174
Scale	1.0437
Weibull Shape	0.9581

Then we have Equation 6 and 7.

$$\hat{\lambda} = exp\left(\frac{-\mu}{\sigma}\right) = exp\left(\frac{-4.1868}{1.0437}\right) = 0.00235 \qquad (6)$$

$$\hat{\gamma} = \frac{1}{\sigma} = \frac{1}{1.0437} = 0.958129 \qquad (7)$$

Based on SAS output, the new cox regression model regresi cox is in Equation 8.

$$h(t) = 0.017347(t)^{-0.0418}exp(-2.1306X_8 - 1.7314X_{9.1}) \qquad (8)$$

To test the significance of the best model, a likelihood ratio test will be conducted. With the help of SAS software, a value of χ^2_{LR} was obtained for 16,25816. Then by using a significance level of 0.05 and degree of freedom 2 a value of $\chi^2_{0.05,2}$ was obtained from the Chi-square table of 5.99. Because $\chi^2_{LR} > \chi^2_{0.05,2}$, then reject H$_0$. This means that the model is significant because there is at least one parameter estimator that is not equal to zero.

Furthermore, for the Wald test, based on SAS output, it is known that the p-value for X$_8$ is 0.0164 and X$_{9.1}$ is 0.0230. By using a significance level of 0.05, the p-values of X$_8$ and X$_{9.1}$ are smaller than 0.05. Thus it can be concluded that the two variables have a significant influence.

Discussion. Based on the analysis results above, the best cox regression model is obtained as Equation 9.

$$h(t) = 0.0025(t)^{-0.0418}exp(-2.1306X_8 + 1.731X_{9.1}) \qquad (9)$$

The Cox regression model is a Cox regression model with a variable X_8 valued at 0 when gestational age <37 weeks and variable X_9.1 valued 0 when the baby's birth weight ≤1499 gr. Then the Cox regression model for the variable X_8 is 1 when gestational age ≥37 weeks and the variable $X_{9.1}$ is 1 when the birth weight of infants 1500-2499 gr is $h(t) = 0.00225(t)^{-0.0418}$.

Based on the likelihood ratio test, it is known that the best cox regression model is significant because it can explain the relationship between the independent variables with the dependent variable. Each parameter in the model is also significant because based on the Wald test it is known that X_8 (gestational age) and $X_{9.1}$ (infant birth weight) have a significant influence on patient survival time.

To determine the probability of LBW risk in Padang Panjang Regional Hospital based on the factors that influence the risk of death, it can be determined through the hazard ratio value of each significant variable. The hazard ratio value are described in Table 4.

Table 4. Hazard ratio value on gestation age < 37 weeks and baby birth's weight $1500 - 2499$ gr

Variable	DF	Estimate $(\hat{\beta})$	Hazard ratio $e^{\hat{\beta}}$
Gestation age (0)	1	-2.1306	0.1187
Baby's birth weight (0)	1	-1.7314	0.1770

Based on table 4, the risk of LBW death for LBW who has a gestational age <37 weeks is 1 / 0.1187 or 8,424 times greater than LBW who has a gestational age ≥of 37 weeks. Then the chance of death risk for LBW with birth weight ≤1499 gr is equal to 1 / 0.1170 or 8,547 times greater than LBW with birth weight 1500-2499 gr. So it can be concluded that the factors that influence the risk of LBW death are gestational age and infant birth weight.

So it can be concluded that the factors that influence the risk of LBW death are gestational age and infant birth weight. Where LBW with gestational age <37 weeks and birth weight <1499 gr greater chance of risk of death compared with LBW with gestational age ≥37 weeks and birth weight 1500-2499 gr.

4   Conclusion

Based on the results of research and discussion, we obtain:
1. The Cox regression model with the best AIC model selection criteria is stated in Equation 10.

$$h(t) = 0.0025(t)^{-0.0418}\exp(-2.1306X_8 + 1.731X_{9.1}) \qquad (10)$$

   Where h (t) is a dense function of opportunity from hazard function or risk of LBW death at the t-time.
2. Factors that influence the risk of death are gestational age and infant birth weight.

References

[1] Maryunani, A.: Asuhan Bayi dengan Berat Badan Lahir Rendah. CV.Trans Info Media. Jakarta (2013)

[2] Cunningham, G., Gutmenti, L.: Obstetri Williams Edisi 23. EGC. Jakarta (2013)

[3] Latief A., Napitulu, P. M., Pudjiadi, A., Ghazali, M. V., Putra, S. T.: Ilmu Kesehatan Anak. FKUI. Jakarta (1985)

[4] Collet,D.: Modelling Survival Data In Medical Research. 2nd edition. Chapman and Hall. London (2003)

[5] Stephen,W.J.: What is A Cox Model. Hayward (2009)